ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Discrete Applied Mathematics 127 (2003) 199–220

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

The k-weak hierarchical representations: an extension of the indexed closed weak hierarchies

P. Bertrand[a,*], M.F. Janowitz[b]

[a] Université Paris-Dauphine, Ceremade, Place du Maréchal De Lattre de Tassigny, 75 775 Paris, Cedex 16, France

[b] DIMACS, Rutgers, The State University, 96 Frelinghuysen Road, Piscataway, NJ 08854, USA

Received 4 March 1999; received in revised form 14 March 2001; accepted 27 August 2001

Abstract

Several approaches have been proposed for the purpose of proving that different classes of dissimilarities (e.g. ultrametrics) can be represented by certain types of stratified clusterings which are easily visualized (e.g. indexed hierarchies). These approaches differ in the choice of the clusters that are used to represent a dissimilarity coefficient. More precisely, the clusters may be defined as the maximal linked subsets, also called M_L-sets; equally they may be defined as a particular type of 2-ball. In this paper, we first introduce the notion of a k-ball, thereby extending the notion of a 2-ball. For an arbitrary dissimilarity coefficient, we establish some properties of the k-balls that pinpoint the connection between them and the M_L-sets. We also introduce the $(2,k)$-point condition ($k \geqslant 1$) which is an extension of the Bandelt four-point condition.

For $k \geqslant 2$, we prove that the dissimilarities satisfying the $(2,k)$-point condition are in one–one correspondence with a class of stratified clusterings, called k-weak hierarchical representations, whose main characteristic is that the intersection of $(k+1)$ arbitrary clusters may be reduced to the intersection of some k of these clusters.

Keywords: k-ball; Maximal linked subset; Weak hierarchy of breath at most k; Extension of the Bandelt four-point condition

1. Introduction

Within the realm of the general area called exploratory data analysis, there is a collection of techniques that together form a discipline called cluster analysis. There is

* Corresponding author. Action Axis, INRIA, Rocquencourt Domaine de voluceau, BP 105, 78 153 Le Chesnay Cedex, France.

E-mail address: patrice.bertrand@inria.fr (P. Bertrand).

0166-218X/03/$ - see front matter
PII: S0166-218X(02)00206-8

a vast literature on this subject—so vast that we can make no attempt to even describe it. Instead we refer the reader to three recent references where a guide to the cluster analysis literature can be found: [1,2,21]. The aim of cluster analysis is to start with a finite set of objects and somehow group them into subsets (called clusters) in such a way that objects in the same subset are in some sense more similar than objects in different subsets. The output of a clustering algorithm may even consist of a nested sequence of such groupings. Unfortunately, there is no agreement as to what should constitute a cluster or as to how clusters should mutually interact. Our goal will be two-fold. First, we compare properties of various types of clusters with a view to understanding how they relate and how they interact. Then we investigate clustering structures and how they relate to certain types of measures of dissimilarity.

We take as our input a dissimilarity coefficient (DC) on the finite set E of objects, and as our output some sort of indexed system of clusters. The types of clustering structures we consider include hierarchies, weak hierarchies, weak hierarchies of breadth at most k (see [5]) and similar structures. The clusters themselves might be linked subsets or maximal linked subsets associated with an appropriate similarity relation, or they might be some sort of generalized ball associated with a DC. The whole point to a cluster algorithm is to obtain an output that has a more easily understandable interpretation than did the input. For that reason we shall investigate properties of hierarchies, weak hierarchies, and generalizations of these ideas. All of this will be made more precise in the next section, where the technical details will be introduced.

We should stress the fact that the clusters that are generated from an underlying DC are rather special subsets. In general, clusters can arise from notions of t-variate dissimilarity, in other words from notions of a type of function defined on t-element subsets, for which $t=2$ is the special case of dissimilarities. Our emphasis will be on $t=2$, and in this case we will show a way to obtain clustering structures that are very similar to those obtained, for $t \geqslant 2$, by Bandelt and Dress [5] and by Diatta [12]. With this aim, we introduce the $(2,k)$-point condition ($k \geqslant 1$) which is an extension of the Bandelt four-point condition defined for any DC. Moreover, we define a class of stratified clusterings, that we call k-weak hierarchical representations, whose main characteristic is that the intersection of $(k+1)$ arbitrary clusters may be reduced to the intersection of some k of these clusters. When $k \geqslant 2$, we prove that the dissimilarities satisfying the $(2,k)$-point condition are in one–one correspondence with the k-weak hierarchical representations, thus providing a bijection somewhat different from those obtained by Bandelt and Dress [5] and by Diatta [12].

In connection with these ideas, we shall consider closure operators defined on the subsets of E having cardinality at least k. These are simply order preserving set mappings ϕ having the property that $A \subseteq \phi(A) = \phi^2(A)$. The image of such a mapping specifies the resulting family of closed sets; it is clearly closed under the formation of intersections having cardinality at least k.

2. Background

We begin with some notations and definitions that are relevant to the theory of cluster analysis. In the following, the set of objects to be clustered, which is denoted by E, is assumed to be nonempty and finite. A mapping $\delta : E \times E \mapsto [0,\infty)$ is said to be a DC defined on E if $\delta(a,b) = \delta(b,a) \geqslant \delta(a,a) = 0$ for all $a,b \in E$. Various classes of DCs are of interest in cluster analysis. For example, it is often assumed that $\delta(a,b) = 0$ implies $a = b$ and, in this case δ is called *proper*. If $\delta(a,b) = 0$ implies $\delta(a,c) = \delta(b,c)$ for all $c \in E$, then δ is said to be *semiproper*. A DC δ is said to satisfy the *ultrametric inequality* if $\delta(a,b) \leqslant \max\{\delta(a,c),\ \delta(c,b)\}$ for all $a,b,c \in E$, and in this case δ is said to be an *ultrametric*. A DC δ is said to satisfy the *Bandelt four-point condition* (see [3,5,14]), if $\delta(a,b) \leqslant \max\{\delta(b,c_1), \delta(c_1,c_2), \delta(c_2,b)\}$ for all $b \in E$ and for all $a, c_1, c_2 \in E$ such that $\max\{\delta(a,c_1), \delta(a,c_2)\} \leqslant \delta(c_1,c_2)$. The Bandelt four-point condition is an extension of the ultrametric inequality. In the sequel, we will introduce a general formalism that leads to an extension of the Bandelt four-point condition.

On the other hand, various classes of collections of subsets have been investigated in cluster analysis. More precisely, let $\mathscr{S}$ be a collection of nonempty subsets (called clusters) of E. The collection $\mathscr{S}$ is called a *partition* if the subsets in $\mathscr{S}$ are disjoint, nonempty and if their union is E. The collection $\mathscr{S}$ is said to be a *covering* if the subsets in $\mathscr{S}$ are noncomparable (hence nonempty) and if their union is E. The collection $\mathscr{S}$ is said to be a *hierarchy on E* if the set E and all the singletons of E belong to $\mathscr{S}$, and $A \cap B \in \{\emptyset, A, B\}$ for all $A, B \in \mathscr{S}$. The collection $\mathscr{S}$ is said to be a *weak hierarchy* if the set $E \in \mathscr{S}$ and $A \cap B \cap C \in \{A \cap B, A \cap C, B \cap C\}$ for all $A, B, C \in \mathscr{S}$ (see [4]). Here it is stipulated that the minimal elements of a weak hierarchy (with respect to the inclusion order) cover the set E (see [3] or [13]). A weak hierarchy is said to be *closed* if it is closed under the formation of arbitrary nonempty intersections. A map $f : \mathscr{S} \mapsto [0,\infty)$ is called an *index* on a weak hierarchy $\mathscr{S}$ if $A \subset B$ implies $f(A) < f(B)$ for all $A, B \in \mathscr{S}$, and if $f(A) = 0$ for any minimal cluster A of $\mathscr{S}$ (with respect to the inclusion order). An *indexed weak hierarchy* will designate any pair $(\mathscr{S}, f)$ where $\mathscr{S}$ is a weak hierarchy, and f is an index on $\mathscr{S}$.

The indexed closed weak hierarchies are in one–one correspondence with the dissimilarities satisfying the Bandelt four-point condition (see [3,13]). Different one–one correspondences exist between classes of indexed closed weak hierarchies, and classes of dissimilarities satisfying the Bandelt four-point condition. These results have been established in distinct ways and by taking into account properties of different types of subsets associated with any dissimilarity, namely the *maximal linked subsets*, the *balls* and the 2-*balls* (see, for example, [7,8,10,11,13,15,16,18–20]).

Let us first recall the definitions of these elementary types of subsets. Given an arbitrary DC δ defined on E, a *maximal linked subset* at level $h \in [0,\infty)$ (see for example [19]) is defined as a subset $M \subseteq E$ that is *linked at level h* in the sense that $a, b \in M$ implies $\delta(a,b) \leqslant h$, and which is maximal with respect to this property. A maximal linked subset is often called an M_{L}-*set*. Let us now consider the map $T\delta$ from $[0,\infty)$ to the set $\Sigma(E)$ of reflexive symmetric relations on E, which is defined by $T\delta(h) = \{(a,b)\colon \delta(a,b) \leqslant h\}$. The M_{L}-sets may then equally be thought of as the *maximal cliques* or the *maximal complete subgraphs* of the threshold graphs associated with the family $\{T\delta(h)\colon h \geqslant 0\}$ of reflexive symmetric relations. So, we will denote by $M_{\mathrm{L}}(T\delta)$ the set of all M_{L}-sets associated with the DC δ.

The *diameter* of any nonempty subset A of E, denoted as $\operatorname{diam}_\delta(A)$ or $\operatorname{diam} A$ if there is no risk of confusion, is defined by $\operatorname{diam} A = \max\{\delta(x,y)\colon x,y \in A\}$. For each nonempty subset A of E, we denote by $\mathscr{M}(A)$ the set of maximal linked subsets that contain A at level $\operatorname{diam} A$; i.e.,

$$\mathscr{M}(A) = \{M \in M_{\mathrm{L}}(T\delta)\colon A \subseteq M, \operatorname{diam} M = \operatorname{diam} A\}.$$

When $A = \{a,b\}$, we will write $\mathscr{M}(a,b)$ instead of $\mathscr{M}(\{a,b\})$. The *ball* of center $a \in E$ and radius $h \in [0,\infty)$, denoted by $\mathbf{B}_\delta(a,h)$, is defined by

$$\mathbf{B}_\delta(a,h) = \{x \in E\colon \delta(x,a) \leqslant h\}.$$

The 2-*ball generated by* $a,b \in E$ (cf., for example, [17] or [13]) is denoted $\mathbf{B}^\delta_{ab}$ and defined as follows:

$$\mathbf{B}^\delta_{ab} = \mathbf{B}_\delta(a,\delta(a,b)) \cap \mathbf{B}_\delta(b,\delta(a,b)). \tag{1}$$

When there is no danger of confusion, the notation $\mathbf{B}_{ab}$ is used in place of $\mathbf{B}^\delta_{ab}$. It is straightforward to show that $\mathbf{B}_{ab} = \{x \in E\colon \max\{\delta(a,x),\delta(b,x)\} \leqslant \delta(a,b)\}$. It is important to note that in the definition of $\mathbf{B}_{ab}$, we are not requiring that $a \neq b$.

We now introduce some general terminology which extends the usual terminology described in the beginning of this section.

Notation 1. Let k be a positive integer. We will write $[k]$ in order to designate the interval $[1,k]$ of the set of natural numbers; i.e., $[k] = \{1,2,\ldots,k\}$. Moreover, for any set X and for any $k \leqslant |X|$, we will denote by $X^{(k)}$ the set of k-element subsets of X, and by $\mathscr{P}_k(X)$ the collection of subsets of X having a cardinality $\geqslant k$.

Definition 1. In the following, a k-*set system* of E will designate any subcollection of $\mathscr{P}_k(E)$ having E as one of its elements.

A 1-set system will be simply called a *set system*. In other words, a set system will be any collection of nonempty subsets of E having E as one of its elements.

A k-*closure system* will designate any k-set system which is closed under intersections having cardinality $\geqslant k$. Following the standard terminology used for cluster analysis, a *1-closure system* will also be called a *closed* set system.

Remark 1. We may notice that any k-closure system $\mathscr{S}$ coincides with the collection of fixed points of the closure operator $\phi_{\mathscr{S}}$ defined on $\mathscr{P}_k(E)$ by $\phi_{\mathscr{S}}(A) = \bigcap\{C \in \mathscr{S}\colon A \subseteq C\}$.

This motivates the following definition:

Definition 2. Any subset of E belonging to a k-closure system $\mathscr{S}$ is said to be $\mathscr{S}$-*closed*.

Notation 2. Given any nonempty subset A of E, we denote

$$\mathbf{B}_A^\delta = \{x \in E\colon \forall a \in A,\ \delta(a,x) \leqslant \operatorname{diam} A\}.$$

It should be noticed that $\mathbf{B}_A^\delta = \bigcap_{a\in A} \mathbf{B}_\delta(a, \operatorname{diam} A)$, and therefore $\mathbf{B}_A^\delta$ coincides with the intersection of exactly $|A|$ balls. When there is no danger of confusion, the notation $\mathbf{B}_A$ will be used in place of $\mathbf{B}_A^\delta$.

Remark 2. Let us consider the following example where δ is a dissimilarity defined on $E = \{a,b,c,d,e\}$ by

	b	c	d	e	f
a	4	3	3	3	5
b		3	3	3	5
c			2	4	2
d				2	5
e					2

Here, $\mathbf{B}_{\{a,b\}} = \mathbf{B}_{\{c,d,e\}} = \{a,b,c,d,e\}$, and consequently the two subsets $A = \{a,b\}$ and $A' = \{c,d,e\}$ satisfy

$$\mathbf{B}_A = \mathbf{B}_{A'}, \qquad |A| \neq |A'| \quad \text{and} \quad A \cap A' = \emptyset.$$

Moreover, $\mathbf{B}_{\{c,e\}} = E$; $\mathbf{B}_{\{d,e\}} = \{d,e\}$ and $\mathbf{B}_{\{d,c\}} = \{d,c\}$. Therefore, despite the fact that $\mathbf{B}_A = \mathbf{B}_{A'}$ with $|A| \neq |A'|$, there is no subset $C \subseteq A'$ such that $\mathbf{B}_C = \mathbf{B}_A$ with $|C| = |A| = 2$. Hence, given any subset $A \subseteq E$ and an arbitrary DC δ, the condition $\mathbf{B}_X = \mathbf{B}_A$ does not imply any connection between X and A, other than the obvious fact that $X \subseteq \mathbf{B}_A$.

Taking this remark into account, we now define a *k-ball*:

Definition 3. The subset $\mathbf{B}_A$ will be called a *k-ball* when $|A| = k$.

Of course a k-ball might also be a k'-ball with $k' \neq k$, but this will not cause any difficulty.

Remark 3. Here, we must mention that Definition 3 with $k = 2$ contradicts slightly the definition of a 2-ball $\mathbf{B}_{ab}$ as defined in Eq. (1). More precisely, the 2-ball $\mathbf{B}_{ab}$ generated by $a, b \in E$, is a 2-ball in the sense of Definition 3, if and only if $a \neq b$, and in this case we have $\mathbf{B}_{ab} = \mathbf{B}_{\{a,b\}}$. Otherwise $a = b$, and $\mathbf{B}_{ab} = \mathbf{B}_{\{a\}}$ is a 1-ball in the sense of Definition 3. In the rest of this text, the term 2-ball will be used in the sense of Definition 3.

Notation 3. For any subset A of E and any element x in E, we will denote Ax as the subset of E which is defined by $Ax = A \cup \{x\}$.

Using Notations 2 and 3, the next result is obvious.

Lemma 1. *Let A and A' be two nonempty subsets of E, and let x be an element of E. Then the following properties hold*:
(1) $(\operatorname{diam} A = \operatorname{diam} A'$ *and* $A \subseteq A') \Rightarrow \mathbf{B}_{A'} \subseteq \mathbf{B}_A$;
(2) $(\operatorname{diam} A = \operatorname{diam} A'$ *and* $\mathbf{B}_A \subseteq A') \Rightarrow \mathbf{B}_{A'} = \mathbf{B}_A = A'$;
(3) $\operatorname{diam} A = \operatorname{diam} Ax \Leftrightarrow x \in \mathbf{B}_A \Leftrightarrow \mathbf{B}_{Ax} \subseteq \mathbf{B}_A$.

Lemma 2. *Let k, k' be two integers such that $k' \geqslant k > 2$ and $A \in E^{(k')}$. If a_1 and a_2 are two distinct elements of A such that* $\operatorname{diam} A = \delta(a_1, a_2)$, *then*

$$\mathbf{B}_A = \bigcap \{\mathbf{B}_Y\colon \{a_1, a_2\} \subset Y, Y \in A^{(k)}\}.$$

Proof. It is obvious from the definition of $\mathbf{B}_A = \bigcap_{a \in A} \mathbf{B}(a, \operatorname{diam} A)$, that

$$\mathbf{B}_A = \bigcap_{Y \in \mathscr{F}_{k;a_1,a_2}(A)} \{u \in E\colon \delta(u, y) \leqslant \delta(a_1, a_2) \quad \text{for all } y \in Y\},$$

where $\mathscr{F}_{k;a_1,a_2}(A) = \{Y \in A^{(k)}\colon a_1, a_2 \in Y\}$, and consequently the k'-ball $\mathbf{B}_A$ may be written as $\mathbf{B}_A = \bigcap_{Y \in \mathscr{F}_{k;a_1,a_2}(A)} \mathbf{B}_Y$, which establishes the lemma. □

Restating Proposition 1 of Bertrand and Janowitz [10], we have

Proposition 1. *Let δ be a DC on E, and let A be any nonempty subset of E. Then* (i) *and* (ii) *are satisfied*:
(i) $\mathbf{B}_A = \bigcup \mathscr{M}(A)$;
(ii) $\operatorname{diam} \mathbf{B}_A = \operatorname{diam} A \Leftrightarrow \mathbf{B}_A \in \mathscr{M}(A) \Leftrightarrow |\mathscr{M}(A)| = 1$.

Remark 4. Another view of the links between M_{L}-sets and k-balls can be provided by the easily proved assertion that the M_{L}-sets coincide with the fixed points of the mapping $X \mapsto \mathbf{B}_X$.

Theorem 1. *If A and C are two nonempty subsets of E, then the following three conditions are equivalent*:
(i) $\operatorname{diam} C = \operatorname{diam}(A \cup C)$,
(ii) $A \subseteq \mathbf{B}_C$ *and* $\operatorname{diam} A \leqslant \operatorname{diam} C$,
(iii) $\mathbf{B}_{A \cup C} \subseteq \mathbf{B}_C$ *and* $\operatorname{diam} A \leqslant \operatorname{diam} C$.

Proof. (iii) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (i) are obvious, so we only have to establish (i) $\Rightarrow$ (iii). Assuming that $\operatorname{diam} C = \operatorname{diam}(A \cup C)$, it is clear that $\operatorname{diam} A \leqslant \operatorname{diam} C$. It is now immediate from Lemma 1(1) that $\mathbf{B}_{A \cup C} \subseteq \mathbf{B}_C$. □

Notation 4. Given two arbitrary subsets A and C of E, let us denote

$$\tilde{\delta}(A, C) = \max\{\delta(a, c)\colon a \in A, c \in C\}.$$

Using this notation, it may be noticed that $A \subseteq \mathbf{B}_C \Leftrightarrow \tilde{\delta}(A, C) \leqslant \operatorname{diam} C$. We observe also that $\operatorname{diam}(A \cup C) = \max\{\operatorname{diam} A,\ \operatorname{diam} C,\ \tilde{\delta}(A, C)\}$.

Using the general terminology concerning a k-closure system (see the definitions previously introduced in this section), we now focus on particular types of subsets of E.

Definition 4. Let $k \geqslant 1$ be an integer and let $\mathscr{S}_k$ denote the k-set system defined by

$$\mathscr{S}_k = \{C \in \mathscr{P}_k(E)\text{: for any } A \in C^{(k)},\ \mathbf{B}_A \subseteq C\}.$$

It is straightforward to prove that the k-set system $\mathscr{S}_k$ is a k-closure system. The $\mathscr{S}_k$-closed subsets will simply be called *k-closed sets*.

Remark 5. It is easily seen that a subset $C \in \mathscr{P}_k(E)$ is k-closed if and only if diam $A < \tilde{\delta}(\{x\}, A)$ for all $x \notin C$ and for all $A \in C^{(k)}$. As was already pointed out by Diatta and Fichet [13], this shows that a nonempty subset is a *weak cluster* (in the sense of Bandelt and Dress [4]) if and only if it is both 1-closed and 2-closed.

Lemma 3. *A dissimilarity is semiproper if and only if each* 2*-ball is* 1*-closed.*

Proof. Assume a DC δ is semiproper. Let u, v be two distinct elements of E, and let $a \in \mathbf{B}_{\{u,v\}}$ and $x \in \mathbf{B}_{\{a\}}$. From $\delta(x,a) = 0$, we deduce that $\delta(x,u) = \delta(a,u)$ and $\delta(x,v) = \delta(a,v)$. Therefore $x \in \mathbf{B}_{\{u,v\}}$ and consequently $\mathbf{B}_{\{a\}} \subseteq \mathbf{B}_{\{u,v\}}$, which proves that each 2-ball is 1-closed. Conversely, assume that each 2-ball is 1-closed, and consider $x, y \in E$ satisfying $\delta(x,y) = 0$. Therefore $x \in \mathbf{B}_{\{y\}} \subseteq \mathbf{B}_{\{y,z\}}$. Then $\delta(x,z) \leqslant \delta(y,z)$. By symmetry in x and y, we have also $\delta(y,z) \leqslant \delta(x,z)$, and so δ is semiproper. □

We will also investigate a collection of subsets of E larger than the collection of k-closed sets—namely the k-diameter subsets.

Definition 5. Given an integer $k \geqslant 2$, a subset $C \in \mathscr{P}_k(E)$ will be called a *k-diameter set* if it satisfies

$$\text{For any } A \in C^{(k)}, \quad \operatorname{diam} C = \operatorname{diam} A \Rightarrow \mathbf{B}_A = C.$$

Note that, by Lemma 1(2), this condition is equivalent to

$$\text{For any } A \in C^{(k)}, \quad \operatorname{diam} C = \operatorname{diam} A \Rightarrow \mathbf{B}_A \subseteq C.$$

Remark 6. Using Definitions 4 and 5, it is easily proved that given any integer $k \geqslant 2$, every k-closed set is a k-diameter set, and every k-diameter set is an M_{L}-set. Therefore, for $k \geqslant 2$, we have

$$\text{(I)} \quad \{k\text{-closed sets}\} \subseteq \{k\text{-diameter sets}\} \subseteq \{k\text{-balls}\} \cap \{M_{\mathrm{L}}\text{-sets}\}.$$

These two inclusion relations are generally strict. In the following examples, namely Examples 1 and 2, we present two DCs for which the first, respectively second, inclusion in (I) holds strictly. But it should be noticed that the main part of this paper is devoted to the particular class of DCs such that these two inclusions are equalities (see Remark 14).

Example 1. Let $E=\{a_1,\ldots,a_k,b,c\}$ where k denotes any integer such that $k\geqslant 2$, and let δ be defined on E by

$$\delta(x,y)=\begin{cases}0 & \text{if } x=y,\\ 2 & \text{if } \{x,y\}=\{a_1,a_k\},\\ 3 & \text{if } \{x,y\}=\{a_1,c\},\\ 1 & \text{otherwise.}\end{cases}$$

For $A=\{a_1,\ldots,a_k\}$, we have $\mathbf{B}_A=A\cup\{b\}$ and $\operatorname{diam}\mathbf{B}_A=2$. It may be noticed that a subset $X\subseteq\mathbf{B}_A$ satisfies $\operatorname{diam}X=\operatorname{diam}\mathbf{B}_A=2$ if and only if $\{a_1,a_k\}\subseteq X\subseteq A\cup\{b\}$. Therefore, if such a subset X has cardinality k, then either $X=A$ or $X=A\cup\{b\}\backslash\{a_i\}$ for some $i\in[1,k]\backslash\{1,k\}$. Having denoted $A[i]=(A\cup\{b\})\backslash\{a_i\}$, then $\mathbf{B}_{A[i]}=A\cup\{b\}=\mathbf{B}_A$ for any $i\in[1,k]\backslash\{1,k\}$, and thus $\mathbf{B}_A$ is a k-diameter set.

Otherwise, $\mathbf{B}_{A[1]}=E\backslash\{a_1\}$, hence $\mathbf{B}_{A[1]}\not\subseteq\mathbf{B}_A$. Thus $\mathbf{B}_A$ is not k-closed. This proves that δ satisfies the first inclusion in (I), in the strict sense.

Example 2. Let $E=\{a_1,\ldots,a_k,b_1,b_2,c_1,\ldots,c_k\}$ where k denotes any integer such that $k\geqslant 2$, and let δ be defined on E by

$$\delta(x,y)=\begin{cases}0 & \text{if } x=y,\\ 2 & \text{if } \{x,y\}=\{b_1,b_2\},\\ 3 & \text{if } \{x,y\}\in\{\{a_i,c_i\}\colon i\in[1,k]\},\\ 1 & \text{otherwise.}\end{cases}$$

Denoting $A=\{a_1,\ldots,a_k\}$, we get $\mathbf{B}_A=A\cup\{b_1,b_2\}$. So the k-ball generated by A is also an M_{L}-set. Let us now consider $C=\{a_1,\ldots,a_{k-2},b_1,b_2\}$ which is also a subset of size k. Then $C\subset\mathbf{B}_A$, $\operatorname{diam}C=2=\operatorname{diam}\mathbf{B}_A$ and $\mathbf{B}_C=A\cup\{b_1,b_2,c_{k-1},c_k\}$. Since $\mathbf{B}_C$ is not contained in $\mathbf{B}_A$, it follows that $\mathbf{B}_A$ is not a k-diameter set, and consequently δ satisfies the second inclusion in (I), in the strict sense.

Lemma 4. *Let $2\leqslant k\leqslant k'\leqslant|E|$. Then every k-closed set in $\mathscr{P}_{k'}(E)$ is also k'-closed, and every k-diameter set in $\mathscr{P}_{k'}(E)$ is also a k'-diameter set.*

Proof. We first consider a k-closed set C, where $k\geqslant 2$. It clearly suffices to prove that C is $(k+1)$-closed whenever $|C|>k$. Let $A\subseteq C$ have cardinality $k+1$, and choose $a,b\in A$ so that $\operatorname{diam}A=\delta(a,b)$. Now choose B so that $\{a,b\}\subseteq B\in A^{(k)}$. Noting that $\operatorname{diam}B=\operatorname{diam}A$, we may apply Lemma 1(1) to see that $\mathbf{B}_A\subseteq\mathbf{B}_B$. Since C is k-closed, $\mathbf{B}_B\subseteq C$, whence $\mathbf{B}_A\subseteq C$, so C is $(k+1)$-closed. The proof for a k-diameter set is similar. $\square$

Remark 7. Let C be a subset such that $|C|\geqslant 2$. It may be deduced from Lemma 4 and Definition 5 that $C=\mathbf{B}_C$ if and only if C is a k-diameter set for some integer k.

3. The k-inclusion and k-diameter conditions

In this section, we will define and discuss two conditions: the k-inclusion condition and the k-diameter condition, where k is any positive integer. Provided that the examined dissimilarity is semiproper, the k-inclusion condition and the k-diameter condition, when holding for both $k = 1$ and 2, characterize, respectively, the inclusion condition and the diameter condition, which were defined by Diatta and Fichet [13,14] in order to characterize the dissimilarities satisfying the Bandelt four-point condition.

Definition 6. Given an integer $k \geqslant 1$, a DC δ on E is said to satisfy the *k-inclusion condition* if

$$\forall A_1, A_2 \in E^{(k)}, \quad A_1 \subseteq \mathbf{B}_{A_2} \Rightarrow \mathbf{B}_{A_1} \subseteq \mathbf{B}_{A_2}.$$

In other words, δ satisfies the k-inclusion condition iff every k-ball is k-closed.

Remark 8. Assume that δ satisfies the 1-inclusion condition. Let $a \in E$ and let b, c be two distinct elements of $\mathbf{B}_{\{a\}}$. We have $\delta(a,b) = \delta(a,c) = 0$. Therefore $a \in \mathbf{B}_{\{b\}}$ and $b \in \mathbf{B}_{\{a\}}$. Consequently $c \in \mathbf{B}_{\{a\}} = \mathbf{B}_{\{b\}}$. We deduce that $\delta(b,c) = 0$. Then $\mathbf{B}_{\{b,c\}} \subseteq \mathbf{B}_{\{b\}} = \mathbf{B}_{\{a\}}$, which proves that each 1-ball is 2-closed. It follows that the 1-inclusion condition implies that each 1-ball is 2-closed.

Remark 9. The inclusion condition was stated by Diatta and Fichet [14, p. 91], as follows: for all $a, b \in E$, $\mathbf{B}_{cd} \subseteq \mathbf{B}_{ab}$, for all $c, d \in \mathbf{B}_{ab}$. Since a and b, as well as c and d, are allowed to be distinct or not, the inclusion condition holds if and only if each 1-ball and each 2-ball are both 1-closed and 2-closed. Using Remark 8 and Lemma 3, we deduce that the inclusion condition holds if and only if δ is semiproper and satisfies both the 1- and 2-inclusion conditions.

Theorem 2. *Let δ be a DC on E and $k \geqslant 2$ be an integer. The first three conditions are then equivalent, and imply each of* (4) *and* (5):

(1) *δ satisfies the k-inclusion condition*;
(2) *the k-balls, the k-closed sets, and the k-diameter sets all coincide*;
(3) *for any $A \in E^{(k)}$, $\mathbf{B}_A$ is the smallest k-ball containing A*;
(4) *every k-ball is an M_{L}-set*;
(5) *the set of k-balls is closed under intersections having cardinality at least k.*

Proof. Taking into account Remark 6, it is immediate that properties (1)–(3) are each equivalent to the assertion that every k-ball is k-closed. Moreover, (2) $\Rightarrow$ (4) is a consequence of Proposition 1(ii), and (2) $\Rightarrow$ (5) is obvious. □

The very weakest condition one could imagine for an informative clustering by means of M_{L}-sets would be that for every subset A of E with $|A| = k$, there be exactly one M_{L}-set containing A at the level $\operatorname{diam} A$. There is a very natural condition equivalent to this property: the *k-diameter condition*.

Definition 7. Given an integer $k \geqslant 1$, a DC δ on E is said to satisfy the *k-diameter condition* if for all $A \in E^{(k)}$, $\operatorname{diam} \mathbf{B}_A = \operatorname{diam} A$.

Remark 10. The diameter condition, defined by Diatta and Fichet [13,14], is clearly equivalent to the k-diameter condition holding for both $k = 1$ and 2.

Theorem 3. *Let δ be a DC on E and $k \geqslant 2$ be an integer. Then the following six conditions are equivalent*:

(1) *δ satisfies the k-diameter condition*;

(2) *for any $A \in E^{(k)}$, $\mathbf{B}_A$ is the unique M_{L}-set containing A at level* $\operatorname{diam} A$;

(3) *for any $A \in E^{(k)}$, there is exactly one M_{L}-set containing A at level* $\operatorname{diam} A$;

(4) *for all $A \in E^{(k)}$ and for all $u, v \in E$,*

$$\max\{\operatorname{diam} Au, \operatorname{diam} Av\} \leqslant \operatorname{diam} A \Rightarrow \delta(u,v) \leqslant \operatorname{diam} A;$$

(5) *every M_{L}-set of cardinality at least k is a k-diameter set*;

(6) *the M_{L}-sets of cardinality at least k coincide with the k-balls*, *and the k-diameter sets.*

Proof. By Proposition 1(ii), it is clear that (2) and (3) are just restatements of (1), and consequently (1) $\Leftrightarrow$ (2) $\Leftrightarrow$ (3). Let us prove (1) $\Leftrightarrow$ (4). Condition (4) is clearly equivalent to the assertion that for every $A \in E^{(k)}$, if $u, v \in \mathbf{B}_A$ then $\delta(u,v) \leqslant \operatorname{diam} A$. But this just says that diam $\mathbf{B}_A = \operatorname{diam} A$ for every $A \in E^{(k)}$, i.e., δ satisfies the k-diameter condition.

(3) $\Rightarrow$ (5): Let M be an M_{L}-set, and let $A \in M^{(k)}$ such that $\operatorname{diam} A = \operatorname{diam} M$. By Proposition 1(i), $M \subseteq \mathbf{B}_A$. Moreover, by using (3) and Proposition 1(ii), we get $\mathbf{B}_A \in \mathscr{M}(A)$, whence $M = \mathbf{B}_A$ and (5) is satisfied.

(5) $\Rightarrow$ (1): Let M be an M_{L}-set, and let $A \in M^{(k)}$ such that $\operatorname{diam} A = \operatorname{diam} M$. By (5), M is a k-diameter set, so $M = \mathbf{B}_A$. Therefore $\operatorname{diam} \mathbf{B}_A = \operatorname{diam} M = \operatorname{diam} A$.

(5) $\Rightarrow$ (6): By (5), every M_{L}-set of cardinality at least k, is a k-diameter set, hence a k-ball. Since (5) $\Leftrightarrow$ (1) $\Leftrightarrow$ (2), every k-ball is an M_{L}-set. Hence, the M_{L}-sets of cardinality at least k, the k-diameter sets, and the k-balls all coincide.

Conditions (6) $\Rightarrow$ (5) is obvious. □

Unlike the case $k \in \{1,2\}$ (see [10]), the next results, and more precisely Proposition 2, show that the k-inclusion condition and the k-diameter conditions are not logically independent for $k \geqslant 3$.

Lemma 5. *Let l', l, k', k be integers such that $l' \geqslant l \geqslant k' \geqslant k \geqslant 2$ and $l > 2$. If every l-ball is k-closed then every l'-ball is k'-closed.*

Proof. From Lemma 4, it is sufficient to prove that every l'-ball is k-closed if every l-ball is k-closed. To this end, it clearly suffices to prove that every $(l+1)$-ball is k-closed whenever every l-ball is k-closed. Let $A \in E^{(l+1)}$, then by using Lemma 2, we get

$$\mathbf{B}_A = \bigcap \{\mathbf{B}_X \colon \{a_1, a_2\} \subset X \in A^{(l)}\}$$

with $\delta(a_1, a_2) = \operatorname{diam} A$. Since every l-ball is k-closed, we deduce that $\mathbf{B}_A$ is also k-closed. □

Corollary 1. *If $k' > k \geqslant 3$ then the k-inclusion condition implies that every k'-ball is k-closed.*

Proof. This follows by using Lemma 5 with $l' = l = k' > k > 2$. □

Proposition 2. *If $k \geqslant 3$ then the k-inclusion condition implies the k-diameter condition.*

Proof. Assume that a DC δ satisfies the k-inclusion condition with $k \geqslant 3$, and let us consider $A \in E^{(k)}$. We aim to prove $\delta(u, v) \leqslant \operatorname{diam} A$ for any $u, v \in \mathbf{B}_A$. We consider only the case $u \notin A$ or $v \notin A$, since otherwise $\delta(u, v) \leqslant \operatorname{diam} A$ is clearly true. Without any loss of generality, we assume $u \notin A$. Then $|Au| = k + 1$, and thus $\mathbf{B}_A \subseteq \mathbf{B}_{Au}$ by Corollary 1. We deduce $v \in \mathbf{B}_{Au}$ and therefore $\delta(u, v) \leqslant \operatorname{diam} Au$. But $\operatorname{diam} Au = \operatorname{diam} A$ since $u \in \mathbf{B}_A$, and consequently $\delta(u, v) \leqslant \operatorname{diam} A$. □

We now investigate some properties that derive from the k-diameter condition or the k-inclusion condition. We first consider a local condition that is related to the k-diameter condition.

Definition 8. Given an integer $k \geqslant 1$, a subset $C \in \mathscr{P}_k(E)$ is said to be *k-bounded* if it satisfies

For any $A \in C^{(k)}$, $\quad \operatorname{diam} \mathbf{B}_A = \operatorname{diam} A$.

Remark 11. By Definition 8, a DC δ satisfies the k-diameter condition iff each k-ball is k-bounded. This is also equivalent to saying that the set E is k-bounded. Otherwise, it could be noticed that if a subset C is not k-bounded then no subset containing C can be k-bounded.

In order to characterize k-bounded sets, we reformulate Theorem 3 to the local point of view introduced by Definition 8. The proof is omitted because it is so similar to the proof of Theorem 3.

Proposition 3. *Let δ be a DC on E and $k \geqslant 1$ be an integer. Then the following conditions are equivalent*:

(1) *C is k-bounded*;

(2) *for any $A \in C^{(k)}$, $\mathbf{B}_A$ is the unique M_{L}-set containing A at level* $\operatorname{diam} A$;

(3) *for any $A \in C^{(k)}$, there is exactly one M_{L}-set containing A at level* $\operatorname{diam} A$;

(4) *condition* (4) *of the Theorem* 3 *for $A \subseteq C$.*

Extending the terminology of Diatta and Fichet [14], we define the *weak k-inclusion condition* as follows.

Definition 9. Let k be any integer such that $k \geqslant 2$. A DC δ on E is said to satisfy the *weak k-inclusion condition* if

$$\forall A, X \in E^{(k)}, \quad (X \subseteq \mathbf{B}_A \text{ and } X \cap A \neq \emptyset) \Rightarrow \mathbf{B}_X \subseteq \mathbf{B}_A.$$

Remark 12. Now recall that the *weak inclusion condition* was defined by Diatta and Fichet [14] as follows.

$$\forall a, b, c \in E, \quad c \in \mathbf{B}_{ab} \Rightarrow \mathbf{B}_{ac} \subseteq \mathbf{B}_{ab}.$$

Therefore the weak 2-inclusion condition is equivalent to the property that the weak inclusion condition holds only for $a, b, c \in E$ such that $c \neq a$ and $a \neq b$.

Proposition 4. *Let k be any integer such that $k \geqslant 2$. If δ satisfies the weak k-inclusion condition, then every k-ball which is k-bounded, is also k-closed.*

Proof. Consider a k-ball $\mathbf{B}_A$ which is k-bounded, and let $X \in \mathbf{B}_A^{(k)}$. If $X \cap A \neq \emptyset$ then $\mathbf{B}_X \subseteq \mathbf{B}_A$, since δ satisfies the weak k-inclusion. We next consider $X \cap A = \emptyset$. Since $\mathbf{B}_A$ is k-bounded, we deduce

$$\operatorname{diam} \mathbf{B}_X = \operatorname{diam} X \leqslant \operatorname{diam} \mathbf{B}_A = \operatorname{diam} A. \tag{2}$$

First, assume that $\operatorname{diam} X \leqslant \tilde{\delta}(A, X) = \max\{\delta(a, x)\colon a \in A,\, x \in X\}$. In this case, we let a' denote an element of A and x' an element of X satisfying $\delta(a', x') = \tilde{\delta}(A, X)$. We also let $X' = X \backslash \{w\}$ where w denotes any element in X which is distinct from x'. Then, using the weak k-inclusion with $a' \in A$ and $X' \subseteq \mathbf{B}_A$, we get

$$\mathbf{B}_{X'a'} \subseteq \mathbf{B}_A. \tag{3}$$

Taking into account the fact that $\operatorname{diam} X \leqslant \tilde{\delta}(A, X) = \delta(a', x')$, we deduce that $\max\{\delta(w, x), \delta(w, a')\colon x \in X'\} \leqslant \delta(a', x') = \operatorname{diam} X'a'$. Hence $w \in \mathbf{B}_{X'a'}$ and, using the weak k-inclusion, $X \cap X'a' = X' \neq \emptyset$, we deduce

$$\mathbf{B}_X \subseteq \mathbf{B}_{X'a'}. \tag{4}$$

Then (3) and (4) give $\mathbf{B}_X \subseteq \mathbf{B}_A$. Let us now examine the second case, i.e., $\operatorname{diam} X > \max\{\delta(a, x)\colon a \in A, x \in X\}$. In this case, $A \subseteq \mathbf{B}_X$, and consequently $\operatorname{diam} X = \operatorname{diam} \mathbf{B}_X \geqslant \operatorname{diam} A$. From (2), we then deduce $\operatorname{diam} X = \operatorname{diam} \mathbf{B}_X = \operatorname{diam} A$. Let us consider $z \in \mathbf{B}_X$. From $A \subseteq \mathbf{B}_X$ and $z \in \mathbf{B}_X$, it follows that $\max\{\delta(z, a)\colon a \in A\} \leqslant \operatorname{diam} \mathbf{B}_X = \operatorname{diam} A$, and then $z \in \mathbf{B}_A$. But this proves $\mathbf{B}_X \subseteq \mathbf{B}_A$, and establishes the proof. $\square$

Example 3. In this example, we aim to prove that the converse of Proposition 4 does not hold, i.e. we aim to prove that the condition "every k-ball which is k-bounded is also k-closed" does not imply the weak k-inclusion condition.

Let $E=\{a_1,\dots,a_k,c,d,e\}$ where k denotes any integer such that $k\geqslant 2$, and let δ be the DC defined on E by

$$\delta(x,y)=\begin{cases}3 & \text{if } \{x,y\}\in\{\{a_1,c\};\ \{c,d\}\},\\ 4 & \text{if } \{x,y\}=\{a_k,d\},\\ 2 & \text{if } \{x,y\}\in\{\{a_i,c\};\ \{a_j,d\}\colon i\neq 1, j\neq k\},\\ 2 & \text{if } \{x,y\}=\{a_1,a_k\},\\ 1 & \text{if } \{x,y\}\in\{\{a_i,e\};\ \{a_j,a_l\}: i,j,l\in[1,k], \{j,l\}\neq\{1,k\}\},\\ 5 & \text{if } \{x,y\}\in\{\{c,e\};\ \{d,e\}\}.\end{cases}$$

Denoting $A=\{a_1,\dots,a_k\}$, $A_i=A\backslash\{a_i\}$ for all $i\in[1,k]$, and more generally $A_I=A\backslash\{a_i\colon i\in I\}$ for all $I\subseteq[1,k]$, the list of all k-balls is given by

$\mathbf{B}_A=Ae$,

$\mathbf{B}_{A_1e}=A_1e$, $\mathbf{B}_{A_ke}=A_ke$, $\mathbf{B}_{A_ie}=Ae$ for all $i\in[2,k-1]$,

$\mathbf{B}_{A_1c}=A_1c$, $\mathbf{B}_{A_kc}=Acd$, $\mathbf{B}_{A_ic}=Ac$ for all $i\in[2,k-1]$,

$\mathbf{B}_{A_1d}=Acd$, $\mathbf{B}_{A_kd}=A_kd$, $\mathbf{B}_{A_id}=Acd$ for all $i\in[2,k-1]$,

$\mathbf{B}_{A_{ij}\cup\{c,d\}}=Acd$ for $k\notin\{i,j\}$, $\mathbf{B}_{A_{ik}\cup\{c,d\}}=A_kcd$ for all $i\neq k$,

$\mathbf{B}_{A_{ij}\cup\{d,e\}}=\mathbf{B}_{A_{ij}\cup\{c,e\}}=\mathbf{B}_{A_{ijl}\cup\{c,d,e\}}=E$ for all i,j,l distinct.

In order to prove that each k-ball, which is k-bounded, is also k-closed, we need to only establish that any k-ball is either k-closed or not k-bounded. We first notice that the k-balls $\mathbf{B}_X$ such that $\operatorname{diam} X=\operatorname{diam} E$, or such that $\mathbf{B}_X=X$, are necessarily k-closed. So it is sufficient to examine the k-balls that do not satisfy either of these two conditions, i.e. the subsets Ae, Ac, Acd and A_kcd.

The k-ball Ae is clearly k-closed. The k-balls Ac, Acd and A_kcd are not k-bounded since they contain the subset A_kc which is of size k and which satisfies

$$\operatorname{diam} A_kc=3 \quad\text{and}\quad diam\,\mathbf{B}_{A_kc}=\operatorname{diam} Acd=4.$$

Therefore, every k-ball which is k-bounded, is also k-closed. Otherwise, for any $i\in[2,k-1]$, we have $a_1\in A_ic$ and $A_1\subset\mathbf{B}_{A_ic}=Ac$. But $\mathbf{B}_{A_1a_1}=\mathbf{B}_A=Ae\nsubseteq\mathbf{B}_{A_ic}=Ac$. Therefore, the weak k-inclusion does not hold, and consequently, the converse of Proposition 4 is not satisfied by δ.

Thus the condition "every k-ball which is k-bounded, is also k-closed", is weaker than the weak k-inclusion condition. Nevertheless, if δ satisfies the k-diameter condition, then each k-ball is k-bounded, and therefore, if we assume moreover the condition that each k-bounded k-ball must be k-closed, we deduce that δ satisfies also the k-inclusion condition. The following result is then an immediate consequence of Proposition 4.

Corollary 2. *Given any integer $k\geqslant 3$, the following conditions are equivalent*:

(i) *δ satisfies the k-inclusion condition*;

(ii) *δ satisfies the k-diameter condition and the weak k-inclusion condition*;

(iii) *δ satisfies the k-diameter condition and each k-ball that is k-bounded is k-closed.*

4. The (2, *k*)-point condition

We now recall that a DC δ satisfies the Bandelt four-point condition (cf. Section 2) if and only if $\delta(a,b) \leqslant \max\{\delta(b,c_1),\delta(c_1,c_2),\delta(c_2,b)\}$ for all $b \in E$ and for all $a,c_1,c_2 \in E$ such that $\max\{\delta(a,c_1),\delta(a,c_2)\} \leqslant \delta(c_1,c_2)$. It is worth noting that this is equivalent to the assertion that $\delta(a,b) \leqslant \operatorname{diam}(Cb)$ for all $b \in E$ and for all $a \in E$ and all $C \in E^{(1)} \cup E^{(2)}$ satisfying $\tilde{\delta}(\{a\},C) \leqslant \operatorname{diam} C$. In this section, we shall study in detail the extension of this assertion to a nonempty subset C of arbitrary size $k \geqslant 1$. Since the k points of such a subset C are symmetrical, but not the points a and b, this extension will be referred to as the $(2,k)$-*point condition*.

Definition 10. Given any integer $k \geqslant 1$, we will say that a DC δ satisfies the $(2,k)$-*point condition* if for all $a \in E$ and all subsets C of size k, we have

$$\tilde{\delta}(\{a\},C) \leqslant \operatorname{diam} C \Rightarrow \text{ for all } b \in E, \quad \delta(a,b) \leqslant \operatorname{diam} Cb.$$

The Bandelt four-point condition is then clearly equivalent to the $(2,k)$-point condition holding for both $k=1$ and 2.

Remark 13. Note that the $(2,1)$-point condition yields $\delta(a,b)=\delta(c,b)$ for all $b \in E$ and for all $a,c \in E$ satisfying $\delta(a,c)=0$. In other words, the (2,1)-point condition is equivalent to the condition that δ is semiproper. The Bandelt four-point condition is therefore equivalent to the condition that δ is semiproper and satisfies the (2,2)-point condition.

The next result introduces two equivalent formulations of the $(2,k)$-point condition for $k \geqslant 1$.

Lemma 6. *For every integer $k \geqslant 1$, the following conditions are equivalent*:
(i) *δ satisfies the $(2,k)$-point condition*;
(ii) *for every $C \in E^{(k)}$ and $x \in E$, $\mathbf{B}_C \subseteq \mathbf{B}_{Cx}$*;
(iii) *for every $C \in E^{(k)}$ and $X \subseteq E$, $\mathbf{B}_C \subseteq \mathbf{B}_{C \cup X}$.*

Proof. Let C be an arbitrary subset of size $k \geqslant 1$, and $x \in E$. The $(2,k)$-point condition may clearly be rewritten as follows:

$$a \in \mathbf{B}_C \Rightarrow \text{for all } x \in E, \quad a \in \mathbf{B}_{Cx}.$$

Hence, the $(2,k)$-point condition is equivalent to (ii). In order to prove that (ii) and (iii) are equivalent, it is sufficient to prove (ii) $\Rightarrow$ (iii). We then suppose (ii), and consider two subsets C and X of E, with $|C|=k$. Since (ii) holds, $\mathbf{B}_C \subseteq \mathbf{B}_{Cx}$ for every $x \in X$. Let $a \in \mathbf{B}_C$, then $\delta(a,x) \leqslant \operatorname{diam} Cx$, for every $x \in X$. Therefore, for every $a \in \mathbf{B}_C$, we have

$$\begin{cases} \text{For all } c \in C, \ \delta(a,c) \leqslant \operatorname{diam} C \leqslant \operatorname{diam}(C \cup X), \\ \text{For all } x \in X, \ \delta(a,x) \leqslant \operatorname{diam} Cx \leqslant \operatorname{diam}(C \cup X). \end{cases}$$

Consequently $\tilde{\delta}(\{a\}, C \cup X) \leqslant \mathrm{diam}(C \cup X)$, which proves $a \in \mathbf{B}_{C \cup X}$. We finally deduce $\mathbf{B}_C \subseteq \mathbf{B}_{C \cup X}$, and then (iii) is satisfied. □

Theorem 4 will show that the condition that each $(k+1)$-ball is k-closed is another characterization of the $(2,k)$-point condition. The next lemma compares several conditions which are extensions of this characterization.

Lemma 7. *Let k and l be integers satisfying $l, k \geqslant 2$, and let us consider the following conditions*:
(1) *each $(k+1)$-ball is k-closed*;
(2) *each l-ball is k-closed*;
(3) *each $(l-1)$-ball of cardinality at least l is k-closed*;
(4) *each M_{L}-set of cardinality at least l is k-closed.*

If $l > k \geqslant 2$ and $l > 3$, then (2) $\Leftrightarrow$ (3) and (1) $\Rightarrow$ (2) $\Rightarrow$ (4). If $l = 3$ and $k = 2$, then (1) and (2) coincide, (2) $\Leftrightarrow$ (4), and (2) strictly implies (3).

Proof. We first assume that $l > k \geqslant 2$ and $l > 3$, and observe that (1) $\Rightarrow$ (2) is then a direct consequence of Lemma 5.

(2) $\Rightarrow$ (3): Let $A \in E^{(l-1)}$ such that $|\mathbf{B}_A| \geqslant l$. So we may choose $u \in \mathbf{B}_A \backslash A$ and a subset X such that $\{a_1, a_2\} \subseteq X \in A^{(k)}$, where a_1, a_2 are two distinct elements of A satisfying $\mathrm{diam}\, A = \delta(a_1, a_2)$. By (2), we have $\mathbf{B}_X \subseteq \mathbf{B}_{Au}$. We have also $\mathbf{B}_A \subseteq \mathbf{B}_X$ since $X \subseteq A$ with $\mathrm{diam}\, X = \mathrm{diam}\, A$, and consequently $\mathbf{B}_A \subseteq \mathbf{B}_{Au}$. On the other hand, $A \subset Au$ with $\mathrm{diam}\, A = \mathrm{diam}\, Au$ since $u \in \mathbf{B}_A$, and thus $\mathbf{B}_{Au} \subseteq \mathbf{B}_A$. We conclude that $\mathbf{B}_A = \mathbf{B}_{Au}$, and so $\mathbf{B}_A$ is k-closed, by using (2).

(3) $\Rightarrow$ (2): Let $A \in E^{(l)}$ and let a_1, a_2 be two distinct elements of A such that $\mathrm{diam}\, A = \delta(a_1, a_2)$. Using Lemma 2 with $l > 3$, we obtain

$$\mathbf{B}_A = \bigcap \{\mathbf{B}_X \colon \{a_1, a_2\} \subset X \in A^{(l-1)}\}.$$

For any subset X that satisfies $X \subset A$ with $\mathrm{diam}\, X = \mathrm{diam}\, A$, we have that $\mathbf{B}_A \subseteq \mathbf{B}_X$, and thus $|\mathbf{B}_X| \geqslant l$. From (3), we then deduce that $\mathbf{B}_A$ is k-closed.

(2) $\Rightarrow$ (4): By Remark 4, each M_{L}-set M is an $|M|$-ball, and therefore is k-closed by applying Lemma 5 when $|M| \geqslant l$.

Assume now that $l = 3$ and $k = 2$. It is obvious that (1) and (2) say the same thing. Implications (2) $\Rightarrow$ (3) and (2) $\Rightarrow$ (4) are proved as in the earlier arguments.

Let us prove (4) $\Rightarrow$ (2). Assuming that (4) holds, Lemma 4 implies that each M_{L}-set of cardinality at least 3, is 3-closed, and by Remark 6, is then a 3-diameter set. Using the equivalence of (5) and (6) in Theorem 3, we deduce that each 3-ball is an M_{L}-set of cardinality at least 3, and therefore is 2-closed by (4). This proves that (2) holds. At last, the example given by Diatta and Fichet [14, p. 91] proves that (2) strictly implies (3). □

Theorem 4. *Let $k \geqslant 2$ be an integer. Then the two following conditions are equivalent*:
(1) *δ satisfies the $(2,k)$-point condition*;
(2) *each $(k+1)$-ball is k-closed.*

If $k \geqslant 3$, *then* (1) *and* (2) *are each equivalent to the following condition*:
(3) δ *satisfies the* k-*inclusion condition.*

Proof. Assume first that $k \geqslant 2$. Lemma 3 shows that (2) $\Rightarrow$ (1) holds. Let us prove now (1) $\Rightarrow$ (2). Given $A \in E^{(k+1)}$ and $X \in \mathbf{B}_A^{(k)}$, we must prove $\mathbf{B}_X \subseteq \mathbf{B}_A$. We denote $A = \{a_1, \ldots, a_k, a_{k+1}\}$ with $\delta(a_k, a_{k+1}) = \operatorname{diam} A$. Since $k \geqslant 2$, we have $a_1 \notin \{a_k, a_{k+1}\}$ and $a_1 \in \mathbf{B}_{A_1}$ where $A_1 = A \backslash \{a_1\}$. By Lemma 1(3) and since $\operatorname{diam} A = \operatorname{diam} A_1$, it follows that $\mathbf{B}_{A_1 a_1} = \mathbf{B}_A \subseteq \mathbf{B}_{A_1}$. But $\mathbf{B}_{A_1} \subseteq \mathbf{B}_{A_1 a_1} = \mathbf{B}_A$ by Lemma 6, and then $\mathbf{B}_A = \mathbf{B}_{A_1}$.

We then have $X \subseteq \mathbf{B}_A = \mathbf{B}_{A_1}$. Let x be an arbitrary element of X. We deduce $X \subseteq \mathbf{B}_{A_1 x}$ by Lemma 6. Therefore $\delta(x', x) \leqslant \operatorname{diam} A_1 x$, for every $x' \in X$. Otherwise, $x \in \mathbf{B}_A = \mathbf{B}_{A_1}$; hence $\max\{\delta(x, a)\colon a \in A_1\} \leqslant \operatorname{diam} A_1$, and then $\operatorname{diam} A_1 x = \operatorname{diam} A_1$. Thus $\delta(x, x') \leqslant \operatorname{diam} A_1$ for every $x, x' \in X$. We then deduce

$$\operatorname{diam} X \leqslant \operatorname{diam} A_1 = \operatorname{diam} A. \tag{5}$$

At last, we consider $y \in \mathbf{B}_X$. For any $a \in A$, we have $y \in \mathbf{B}_X \subseteq \mathbf{B}_{Xa}$ by Lemma 6. Therefore $\delta(y, a) \leqslant \operatorname{diam} Xa$. But, since $X \subseteq \mathbf{B}_A$, we have

$$\max\{\delta(x, a)\colon x \in X, a \in A\} \leqslant \operatorname{diam} A. \tag{6}$$

Using (5) and (6), we deduce $\operatorname{diam} Xa \leqslant \operatorname{diam} A$ for every $a \in A$. Hence $\delta(y, a) \leqslant \operatorname{diam} Xa \leqslant \operatorname{diam} A$. Since a denotes an arbitrary element of A, it follows that $y \in \mathbf{B}_A$, and consequently $\mathbf{B}_X \subseteq \mathbf{B}_A$, which proves finally that (2) holds.

Assume now that $k \geqslant 3$. Then (2) $\Leftrightarrow$ (3) results from the equivalence of conditions (2) and (3) of Lemma 7, when $l = k + 1$ with $k \geqslant 3$. □

Remark 14. It follows from Corollary 2 and Theorem 4 that two other geometric conditions characterize the $(2, k)$-point condition for $k \geqslant 3$: one is provided by the k-diameter condition plus the weak k-inclusion condition. The other one is provided by the k-diameter condition together with each k-ball that is k-bounded, also being k-closed. It should be noticed that when holding for both $k = 1$ and 2, each of these two conditions characterizes the Bandelt four-point condition (see, respectively, [14] and [10]).

Moreover, Theorem 4 together with Proposition 2, and Theorems 2 and 3, imply that a DC δ satisfying the $(2, k)$-point condition ($k \geqslant 3$) is also characterized by the fact that the M_{L}-sets of cardinality at least k coincide with the k-balls, and the k-closed sets. This condition holding for both $k = 1$ and 2, characterizes the Bandelt four-point condition (see Corollary 4 in [10]).

Remark 15. Let us consider an arbitrary dissimilarity. The $(2, |E|)$-point condition clearly holds. Furthermore, for all $C \in E^{(|E|-1)}$, we have $Cx = E$ for all $x \notin C$, and so $\mathbf{B}_C \subseteq \mathbf{B}_{Cx}$ for all $x \in E$. Then, the $(2, |E| - 1)$-point condition also holds. Otherwise, it can be noticed that the $(2, k)$-point condition is getting weaker as k is increasing: this results from (1) $\Leftrightarrow$ (2) in Theorem 4 together with Lemma 5.

5. *k*-Weak hierarchical representations

In this section, we shall determine a class of pairs $(\mathscr{S}, f)$, where $\mathscr{S}$ is a set system and f an isotone map from $\mathscr{S}$ into $[0, \infty)$, which is in one–one correspondence with the class of all the dissimilarities satisfying the $(2, k)$-point condition. This leads to an extension of the bijection obtained by Bandelt [3] and in a different way by Diatta and Fichet [13,14], for dissimilarities satisfying the Bandelt four-point condition.

We first define the notions of *pre-indexed set system* and *weakly indexed set system*, as introduced in [9].

Definition 11. Let $\mathscr{S}$ be a set system on E, and f a mapping from $\mathscr{S}$ into $[0, \infty)$. When f is isotone, in other words when $f(A) \leqslant f(B)$ for any $A, B \in \mathscr{S}$ such that $A \subset B$, the mapping f will be called a *pre-index*, and the pair $(\mathscr{S}, f)$ will be called a *pre-indexed set system*.

A *weakly indexed set system* is a pre-indexed set system $(\mathscr{S}, f)$ such that

$$\forall A, B \in \mathscr{S}, \quad A \subset B \quad \text{with } f(A) = f(B) \Rightarrow A = \bigcap\{C \in S \colon A \subset C\}.$$

Remark 16. The notions of pre-indexed set system and weakly indexed set system extend the notion of an *indexed* clustering structure, in the sense that this notion was defined for the indexed hierarchies and for the indexed weak hierarchies (see, for example, [14]). It should be noticed that this extension does not require that f vanish on the minimal clusters of the structure $\mathscr{S}$, as it was required for the indexed closed weak hierarchies in order to obtain a one–one correspondence between these clustering structures and the dissimilarities satisfying the Bandelt four-point condition (see [3,13,14]).

Definition 12. Let $(\mathscr{S}, f)$ be a pre-indexed set system and ρ the symmetric mapping defined on $E \times E$ by $\rho(x, y) = \min\{f(X) \colon X \in \mathscr{S} \text{ and } x, y \in X\}$. First, it is worth noting that the mapping ρ is well defined, since by definition of a set system, any set system $\mathscr{S}$ contains the whole set E. It will then be said that $(\mathscr{S}, f)$ *induces* the symmetric mapping ρ.

As the set system $M_{\mathrm{L}}(T\delta)$ is generally not closed under nonempty intersections, it will be convenient to let $\widehat{M_{\mathrm{L}}}(T\delta)$ be all nonempty intersections of elements of $M_{\mathrm{L}}(T\delta)$. In the following, we will denote by $\mathscr{D}$ the set of all the DCs on E, and we will focus on the map ψ defined from $\mathscr{D}$ to the set of all the pre-indexed set systems, as follows:

$$\psi(\delta) = (\widehat{M_{\mathrm{L}}}(T\delta), \mathrm{diam}_\delta).$$

Given a closed set system $\mathscr{S}$, for each nonempty subset A of E, we will let $A^\star$ denote the element of $\mathscr{S}$ defined by $A^\star = \bigcap\{X \in S \colon A \subseteq X\}$. Note that if $(\mathscr{S}, f)$ is a pre-indexed closed set system, then

$$\forall x, y \in E, \quad \rho(x, y) = f(\{x, y\}^\star).$$

Moreover, it is easy to prove that if $(\mathscr{S}, f)$ is a pre-indexed set system which satisfies

$$(\mathrm{R}_0) \quad \bigcup\{A \in \mathscr{S} : f(A) = 0\} = E,$$

then ρ is a DC on E. Otherwise, we must consider pseudo-dissimilarities on E. A *pseudo-dissimilarity* is a mapping $\rho : E \times E \mapsto [0, \infty)$ such that
(1) $\rho(x, y) = \rho(y, x)$, and
(2) $\rho(x, y) \geqslant \rho(x, x)$
for all $x, y \in E$. We then consider the mapping $\phi : (\mathscr{S}, f) \mapsto \rho$ where $\rho = \phi((\mathscr{S}, f))$ denotes the pseudo-dissimilarity induced by the pre-indexed set system $(\mathscr{S}, f)$.

Batbedat [6] has proposed a general one–one correspondence between dissimilarities and a class of pre-indexed set systems. We shall follow a different but related approach due to Bertrand [9].

Definition 13. Let $(\mathscr{S}, f)$ be a closed and pre-indexed system, and denote by (G) the property defined by

$$f\left(\left[\bigcup_{1 \leqslant i < j \leqslant 3} (C_i \cap C_j)\right]^{\star}\right) \leqslant \max_{1 \leqslant i \leqslant 3} f(C_i)$$

for all C_1, C_2, C_3 elements of $\mathscr{S}$ that are not pairwise disjoint.

We will let $\mathscr{C}_0$ denote the collection of all weakly indexed closed set systems that satisfy the two conditions (G) and (R_0).

The next result was formulated by Bertrand [9].

Theorem 5. *The mapping ψ is a one–one correspondence from $\mathscr{D}$ onto the collection $\mathscr{C}_0$ of all weakly indexed closed set systems satisfying* (G) *and* (R_0). *The converse of this one–one correspondence is the restriction of ϕ to the collection $\mathscr{C}_0$.*

The next proposition together with its proof, is provided in [9].

Proposition 5. *Let $(\mathscr{S}, f)$ be a pre-indexed closed set system, and ρ the mapping induced by $(\mathscr{S}, f)$. Then the following conditions are equivalent*:
(i) *$(\mathscr{S}, f)$ satisfies* (G);
(ii) *for any subset A of E,* $\mathrm{diam}_\rho A = f(A^{\star})$.

We point out the following result which follows from Proposition 5.

Corollary 3. *Let $(\mathscr{S}, f)$ be a pre-indexed closed set system, and ρ the mapping induced by $(\mathscr{S}, f)$. If $(\mathscr{S}, f)$ satisfies* (G), *then $A^{\star} \subseteq \mathbf{B}_A^{\rho}$ for any subset A of E.*

Proof. Let $x \in A^{\star}$, then $Ax \subseteq A^{\star}$, and consequently $(Ax)^{\star} = A^{\star}$. By Proposition 5, we have $\mathrm{diam}_\rho Ax = \mathrm{diam}_\rho A$, which implies $x \in \mathbf{B}_A^{\rho}$. □

Let us recall that Bandelt and Dress [4] defined a weak hierarchy as being a set system $\mathscr{H}$ such that for all $A_1, A_2, A_3 \in \mathscr{H}$:

$$\text{(W)} \quad A_1 \cap A_2 \cap A_3 = A_i \cap A_j, \quad \text{for some } i, j \in \{1, 2, 3\}.$$

This condition (W) was extended by Bandelt and Dress [5] and by Diatta [12] who considered set systems satisfying the following condition (kW).

Definition 14. Given any integer $k \geqslant 2$, we will say that a set system $\mathscr{S}$ satisfies condition (kW) if for all $A_1, \ldots, A_{k+1} \in \mathscr{S}$, we have

$$(k\mathrm{W}) \quad \bigcap_{i \in [k+1]} A_i \in \left\{ \bigcap_{i \in [k+1] \setminus \{j\}} A_i \colon 1 \leqslant j \leqslant k+1 \right\}.$$

Bandelt and Dress [5] and Diatta [12] investigated set systems satisfying condition (kW) together with the property of being equipped with a strictly isotone pre-index. In each of these papers, the authors established a one–one correspondence between some class of these clustering structures and some class of k-way dissimilarities. In the rest of this text, since we restrict our attention to the 2-way dissimilarities, we will look at set systems satisfying condition (kW) in a different way, that is from the point of view provided by Theorem 5 (see also [9]).

Remark 17. Two conditions equivalent to (W) have been formulated by Bandelt and Dress [4] (cf. also [14]). These two conditions can be extended in order to provide two characterizations of condition (kW). More precisely, it can be easily proved that $k+1$ subsets $A_1, \ldots, A_{k+1} \in \mathscr{S}$ satisfy property (kW) if and only if one of the following equivalent properties holds:

$$(k\mathrm{W}') \quad \bigcup_{j \in [k+1]} \left(\bigcap_{i \in [k+1] \setminus \{j\}} A_i \right) \subseteq A_\alpha, \quad \text{for some } \alpha \in [k+1];$$

(kW$''$) There are no $k+1$ elements $x_1, \ldots, x_{k+1}$ s.t. $x_i \in A_j$ iff $i \neq j$.

Definition 15. Given a pre-indexed set system $(\mathscr{S}, f)$, let (I_k) denote the following condition:

$$(\mathrm{I}_k) \quad \text{For all } A, B \in \mathscr{S} \text{ with } |A| \geqslant k,\ A \subset B \Rightarrow f(A) < f(B).$$

A *k-weak hierarchical representation* will denote any pair $(\mathscr{S}, f)$ in $\mathscr{C}_0$ such that $\mathscr{S}$ satisfies condition (kW) and $(\mathscr{S}, f)$ satisfies the condition (I_k).

Proposition 6. *If a DC δ satisfies the $(2,k)$-point condition with $k \geqslant 2$, then $(\widehat{M_{\mathrm{L}}}(T\delta)$, $\mathrm{diam}_\delta)$ is a k-weak hierarchical representation.*

Proof. From Theorem 5, the pair $(\widehat{M_{\mathrm{L}}}(T\delta), \mathrm{diam}_\delta)$ belongs to $\mathscr{C}_0$. So we need only to prove that $\widehat{M_{\mathrm{L}}}(T\delta)$ satisfies (kW) and that $(\widehat{M_{\mathrm{L}}}(T\delta), \mathrm{diam}_\delta)$ satisfies (I_k). We first consider $k+1$ arbitrary subsets $A_1, \ldots, A_{k+1} \in \widehat{M_{\mathrm{L}}}(T\delta)$, and aim to prove that these subsets satisfy condition (kW). Note that if one of the $A_i's$ has size smaller than k, then (kW) has to be satisfied: see the equivalent condition (kW$''$) in Remark 17. So we assume, without any loss of generality, that all the subsets $A_1,\dots,A_{k+1}$ have size $\geqslant k$. Then, using Theorem 3, we first deduce that each subset A_i for $i=1,\dots,k+1$ is the intersection of some k-balls. Hence, from Theorem 2(5), each subset A_i for $i=1,\dots,k+1$, is a k-ball, and then a k-closed set. Let us assume that $A_1,\dots,A_{k+1}$ do not satisfy condition (kW). Then by Remark 17, there exist elements $x_1,\dots,x_{k+1}\in E$ such that for all $i\in[k+1]$,

$$x_i \in \left(\bigcap_{j\in[k+1]\setminus\{i\}} A_j\right)\Big\backslash A_i\ .$$

Let us now denote $X_i=\{x_j\colon j\in[k+1]\setminus\{i\}\}$, for $i\in[k+1]$. Then $\mathbf{B}_{X_i}\subseteq A_i$ (with $i\in[k+1]$), since the subsets A_i are k-closed and the elements x_i are pairwise distinct. It follows that

$$(*)\quad x_i \in \left(\bigcap_{j\in[k+1]\setminus\{i\}} \mathbf{B}_{X_j}\right)\Big\backslash \mathbf{B}_{X_i}$$

for every index $i=1,\dots,k+1$.

Let x_u,x_v be two elements among $x_1,\dots,x_{k+1}$ such that

$$\delta(x_u,x_v)=\max\{\delta(x_j,x_{j'})\colon j,j'\in[k+1]\}.$$

Since $k\geqslant 2$, there exists some $\alpha\in[k+1]$ for which $\{x_u,x_v\}\subseteq X_\alpha$, and then $x_i\in\mathbf{B}_{X_\alpha}$ for each $i=1,\dots,k+1$, which is contradictory with $(*)$.

We prove now that $(\widehat{M_{\mathrm{L}}}(T\delta),\mathrm{diam}_\delta)$ satisfies (I_k). Let $A,B\in\widehat{M_{\mathrm{L}}}(T\delta)$ such that $A\subset B$ and $|A|\geqslant k$. Using Theorems 3 and 4 together with Theorem 2, we deduce that A and B are M_{L}-sets, and thus if $A\subset B$ it is necessary that $\mathrm{diam}\,A<\mathrm{diam}\,B$. $\square$

Proposition 7. *Let $k\geqslant 2$ be an integer, and let $(\mathscr{S},f)$ be a weakly indexed closed set system such that $\mathscr{S}$ satisfies* (kW) *and $(\mathscr{S},f)$ satisfies* (I_k) *and* (G). *Denoting by ρ the mapping induced by $(\mathscr{S},f)$, we have $\mathbf{B}^\rho_A=A^\star$ for any subset of cardinality at least k.*

Proof. Let A be a subset of cardinality at least k. By Corollary 3, it is sufficient to prove $\mathbf{B}^\rho_A\subseteq A^\star$. We will prove that $x\notin A^\star$ implies that $x\notin\mathbf{B}^\rho_A$, and we will assume $A^\star\neq E$, since otherwise the proof is obvious. Let us then consider $x\notin A^\star$, and denote $A_1,\dots,A_k$ a partition of A formed with k subsets. We then define $C_i=Ax\backslash A_i$ for $i=1,\dots,k$, and note that $C_i^\star$ exists for C_i is not empty (since $x\in C_i$) and $\mathscr{S}$ is a closed set system. By hypothesis, the intersection $I=C_1^\star\cap\cdots\cap C_k^\star\cap A^\star$ is equal to an intersection of k subsets among the $k+1$ subsets $C_1^\star,\dots,C_k^\star,A^\star$. Since $x\notin A^\star$, then $x\notin I$, and $I\neq C_1^\star\cap\cdots\cap C_k^\star$. Therefore $I=A^\star\cap(\bigcap_{i\neq j}C_i)$ for some $j\in[1,k]$. Without any loss of generality, we may then assume

$$I=A^\star\cap C_2^\star\cap\cdots\cap C_k^\star.$$

We deduce that $A_1 \subseteq C_1^\star$, thus $A \subseteq C_1^\star$, and consequently $A^\star \subseteq C_1^\star$. Moreover, this inclusion is strict, since $x \notin A^\star$. It follows that $f(A^\star) < f(C_1^\star)$, and $\mathrm{diam}_\rho A < \mathrm{diam}_\rho C_1 \leqslant \mathrm{diam}_\rho Ax$ by Proposition 5. Therefore $x \notin \mathbf{B}_A^\rho$. □

Theorem 6. *Given any integer $k \geqslant 2$, the mapping ψ defines a one–one correspondence from the set of dissimilarities satisfying the $(2,k)$-point condition onto the set of k-weak hierarchical representations.*

Proof. By Proposition 6, $\psi(\delta)$ is a k-weak hierarchical representation, if δ satisfies the $(2,k)$-point condition. Using Theorem 5, we need only prove that if $(\mathscr{S}, f)$ is a k-weak hierarchical representation, then the dissimilarity $\rho = \phi((\mathscr{S}, f))$, which is induced by $(\mathscr{S}, f)$, satisfies the $(2,k)$-point condition. Let A be an arbitrary subset of size k and v an element of E. By Lemma 6, it is sufficient to prove $\mathbf{B}_A^\rho \subseteq \mathbf{B}_{Av}^\rho$ in order to say that ρ satisfies the $(2,k)$-point condition. Since $A \subseteq Av$, we get $A^\star \subseteq (Av)^\star$, and consequently $\mathbf{B}_A^\rho \subseteq \mathbf{B}_{Av}^\rho$ by Proposition 7, which completes the proof. □

Remark 18. Recall that the Bandelt four-point condition holds if and only if the dissimilarity δ is semiproper and satisfies the $(2,2)$-point condition (see Remark 13). Since $\mathrm{diam}_\delta^{-1}(0)$ is a partition of E whenever δ is semiproper (see [9]), we deduce that $\psi(\delta) = (\widehat{M_{\mathrm{L}}}(T\delta), \mathrm{diam}_\delta)$ is an indexed closed weak hierarchy. It follows that the aforementioned bijection between indexed closed weak hierarchies and dissimilarities satisfying the Bandelt four-point condition, is a restriction of the one–one correspondence provided by Theorem 6 with $k=2$.

Acknowledgements

The authors are grateful to the referees for their careful reading of the paper and especially for their many constructive criticisms and helpful suggestions. The first author was partially supported by INRIA-Rocquencourt. The second author was supported by ONR Grant N-00014-96-1-1201.

References

[1] P. Arabie, L.J. Hubert, Advances in cluster analysis relevant to marketing research, in: W. Gaul, D. Pfeifer (Eds.), From Data to Knowledge: Theoretical and Practical Aspects of Classification, Data Analysis, and Knowledge, Springer, Berlin, 1995, pp. 3–19.

[2] P. Arabie, L.J. Hubert, An overview of combinatorial data analysis, in: P. Arabie, L.J. Hubert, G. De Soete (Eds.), Clustering and Classification, World Scientific Publishers, River Edge, NJ, 1996, pp. 5–63.

[3] H.-J. Bandelt, Four-point characterization of the dissimilarity functions obtained from indexed closed weak hierarchies, Mathematisches Seminar, Universität Hamburg, Germany, 1992.

[4] H.-J. Bandelt, A.W.M. Dress, Weak hierarchies associated with similarity measures: an additive clustering technique Bull. Math. Biol. 51 (1989) 133–166.

[5] H.-J. Bandelt, A.W.M. Dress, An order theoretic framework for overlapping clustering, Discrete Math. 136 (1994) 21–37.

[6] A. Batbedat, Les isomorphismes HTS et HTE (après la bijection de Benzécri-Johnson), Metron 46 (1988) 47–59.
[7] J.-P. Benzécri, L'Analyse des données: la Taxinomie, Vol. 1, Dunod, Paris, 1973.
[8] P. Bertrand, Structural properties of pyramidal clustering, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 19 (1995) 35–53.
[9] P. Bertrand, Set systems and dissimilarities, European J. Combin. 21 (2000) 727–743.
[10] P. Bertrand, M.F. Janowitz, Pyramids and weak hierarchies in the ordinal model for clustering, Discrete Appl. Math. 122 (2002) 55–81.
[11] F. Critchley, B. Van Cutsem, An order-theoretic unification and generalisation of certain fundamental bijections in mathematical classification, LMC-Imag Rapports de recherche, RR 874-M and RR 875-M, France, 1992.
[12] J. Diatta, Dissimilarités multivoies et généralisations d'hypergraphes sans triangles, Math. Inf. Sci. Hum. 138 (1997) 57–73.
[13] J. Diatta, B. Fichet, From Apresjan hierarchies and Bandelt-Dress weak hierarchies to quasi-hierarchies, in: E. Diday, et al., (Eds.), New Approaches in Classification and Data Analysis, Springer, Berlin, 1994, pp. 111–118.
[14] J. Diatta, B. Fichet, Quasi-ultrametrics and their 2-ball hypergraphs, Discrete Math. 192 (1998) 87–102.
[15] E. Diday, Une représentation visuelle des classes empiétantes: les pyramides, Rapport de recherche I.N.R.I.A. No. 291, Rocquencourt, France, 1984.
[16] E. Diday, Orders and overlapping clusters in pyramids, in: J. De Leeuw, et al., (Eds.), Multidimensional Data Analysis, DSWO Press, Leiden, 1986, pp. 201–234.
[17] C. Durand, B. Fichet, One-to-one correspondences in pyramidal representation: a unified approach, in: H.H. Bock (Ed.), Classification and Related Methods of Data Analysis, North-Holland, Amsterdam, 1988, pp. 85–90.
[18] M.F. Janowitz, An order theoretic model for cluster analysis, SIAM J. Appl. Math. 34 (1978) 55–72.
[19] N. Jardine, R. Sibson, Mathematical Taxonomy, Wiley, New York, 1971.
[20] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (1967) 241–254.
[21] B. Mirkin, Mathematical Classification and Clustering, Kluwer, Dordrecht, 1996.